

# 普通名詞換言辞書の構築

長岡技術科学大学 電気系

山形 祐輝

山本 和英

# 背景・目的

言語処理の各分野で換言は用いられる  
換言知識の収集は以下の方法が一般的

- シソーラス、国語辞典の語釈文から収集
- コーパス、WEBの文中の関係から抽出

人による換言の知識が得られるような  
汎用な換言辞書は作られていない



**完全な人手により普通名詞換言辞書を構築**

例)「折り鶴」→「紙で折った鶴」、「肴」→「酒のつまみ」

# 辞書の構築

換言対象は形態素解析器JUMANから抽出  
→形態素辞書に含まれる普通名詞16,524語

- 対象語を作業者の考えだけで言い換える
- 換言は簡易な表現を意識する
- 内容語は2~3語程度に収める  
例)「土嚢」→「土を詰めた袋」
- 多義語でない限り、換言対は1つ

# 辞書の構築 ―作業基準―

## ● 換言は付与されているカテゴリに従う

例) 「クラス」カテゴリ：組織・団体 → 「集団」  
          抽象物 → 「階級」  
「王冠」 カテゴリ：人工物-衣服 → 「王がかぶる飾り」  
          × 「瓶のふた」：カテゴリにそぐわない

## ● 以下の場合は換言しない

- 換言語を思いつかない  
「ストライク」「アルカリ」etc.
- 元の語の意味がわからない  
「村八分」「氏神」etc.



作業の効率を上げる  
無理な換言を行わない

# 辞書の構築 — 作業結果 —

換言対象：16,524語

換言対：**16,153語**

換言していない語：980語

- ・ 換言を思いつかなかった語：310語
- ・ 意味が分からない語：670語

換言対象

≠

換言対数

+

無換言語数

↳ 多義語の場合もあるため

# 辞書の構築 — 作業結果 —

カテゴリ	換言対象	換言作成	無記入
人工物	2,610語	2,557語	72語
自然物	453語	420語	33語
場所	1,795語	1,685語	111語
組織・団体	248語	228語	20語
人	1,479語	1,419語	66語
動物	771語	724語	47語
植物	339語	316語	23語
抽象物	6,912語	6,465語	435語
時間	259語	227語	33語
数量	353語	325語	29語
形・模様	135語	120語	15語
色	88語	84語	4語
複数	825語	1,583語	92語
合計	16,267語	16,153語	980語

# 辞書の構築 — 具体例 —

カテゴリ	換言対象	換言対
抽象物	朝飯前	簡単なこと
人工物-衣類	上履き	屋内用の靴
人	OB	卒業した人
抽象物	OB	ゴルフで場外に打つこと
動物	海綿	D
植物	カビ	N

N:換言を思いつかなかった

D:元の語の意味が分からない

カテゴリ	換言対象	換言対
色	藍色	濃い青
植物	藍	青い花
動物	揚げ羽蝶	蝶
場所-施設部位	明かり取り	天窓
植物	麻	布材料の草
時間	当たり年	都合のいい年
自然物	朝露	露
自然物	朝日	朝上ってくる太陽
場所-その他	朝市	朝に開催される市
数量	値/あたい	数値
人	相棒	コンビニの相手
人	相方	コンビニの相手
人	相手方	相手
場所-機能	相手方	相手の側
人	相手	対する人
色	浅緑	淡い緑
場所-自然	浅瀬/	浅い水たまり
動物-部位	赤毛	赤い毛
人工物-食べ物	赤米	赤いお米
自然物	赤土	赤い土
動物	赤虫	赤い虫
動物-部位	赤身	赤い身

カテゴリ	換言対象	換言対
色	赤色	赤
植物	赤松	針葉樹
人	赤子	赤ちゃん
自然物	赤錆び	赤い錆
動物-部位	赤ら顔	赤い顔
人	赤ちゃん	幼児
色	赤	リンゴの色
植物	青黴	青いカビ
植物-部位	青葉	緑の葉
動物	青虫	芋虫
動物	青大将	蛇
色	青色	青
植物	青菜	葉野菜
植物	青海苔	海藻
人工物-食べ物	青海苔	細かくした海苔
色	青	空の色
場所-施設部位	上り口	入口
人工物-その他	証	証拠
抽象物	証	証明
時間	曙	明け方
時間	十六夜	夏の夜
植物	秋草	秋の草花



# 換言とクエリ拡張

## Ellen Mら[1]の研究

結果が一意に決まらない単語のクエリの拡張  
→ WordNetの同義語、上位語、下位語が有効

換言辞書 ≡ 同義表現を集めた語彙資源  
→ WordNetの同義語と同様、クエリ拡張に有効



構築した換言辞書の有用性をクエリ拡張にて示す

[1]Ellen M、Voorhees. Query Expansion using Lexical-SemanticRelations. In 17th International Conference on Research and development in Information Retrieval (SIGIR'94). p61-69, Springer London, 1994.1.

# 辞書の評価 – 使用データ –

換言辞書：普通名詞換言辞書と用言等換言辞書[2]

※用言等換言辞書：サ変名詞、動詞、形容詞、副詞を手で換言したもの

[2] 山本 和英、吉倉 孝太郎. 用言等換言辞書を人手で作りました. 言語処理学会  
第19回年次大会 発表論文集 (2013.3)、pp.276-279

比較対象：日本語WordNet同義語データベース  
ver.1.0

検索対象：毎日新聞2年分(1999、2000)

検索クエリ：換言辞書とWordNetの両方で  
見出し語となっている  
普通名詞とサ変名詞の組み合わせ

# 辞書の評価

検索クエリを各語彙資源で換言して拡張を行い  
元のクエリで得た文と拡張したクエリで得た文の  
内容語で類似度計算を行う

類似度計算はJaccard係数とSimpson係数を用いる

$$Jacc = \frac{|X \cap Y|}{|X \cup Y|} \quad Simp = \frac{|X \cap Y|}{\min(|X|, |Y|)}$$

X: 元クエリで獲得した文の内容語の集合

Y: 拡張して獲得した文の内容語の集合

# 辞書の評価 — 結果 —

	検索クエリ	換言辞書 による拡張	WordNet による拡張
クエリ数	24,510組	+73,530組	+1,074,212組
獲得文数	140,604文	+110,237文	+110,151文



換言辞書は文に出現しやすい語に拡張している

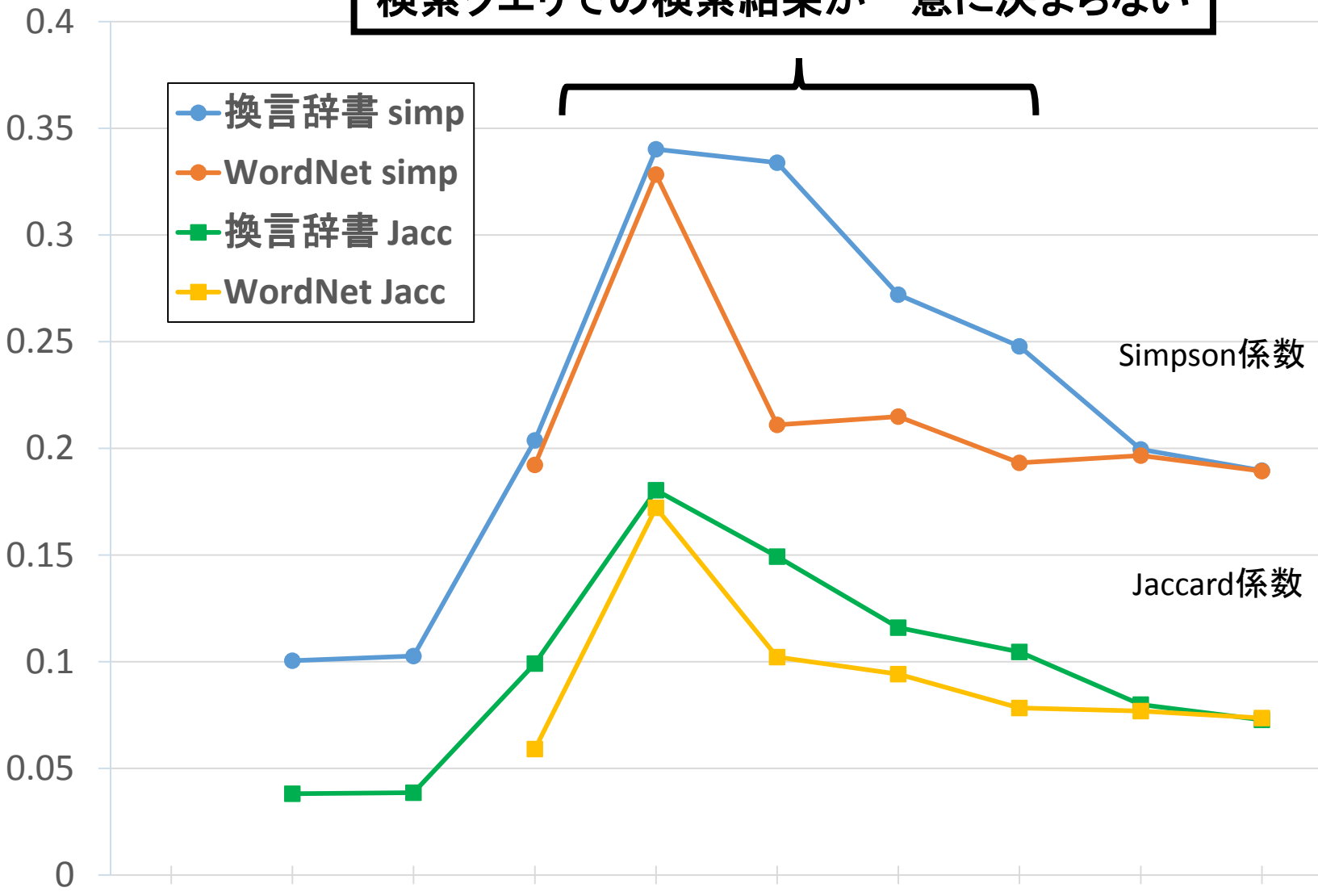
拡張の例) 検索クエリ「負債 削減」  
換言辞書「借金 削減」「負債 減らす」  
「借金 減らす」  
WordNet「借入 削減」「負債 カット」  
「負い目 カット」「金銭債務 削減」  
「借入 カット」 etc.

# 検索クエリでの検索結果が一意に決まらない

検索クエリで獲得した文と拡張で獲得した文  
↑ 近い  
内容  
↓ 違う

Simpson係数  
Jaccard係数

- 換言辞書 simp
- WordNet simp
- 換言辞書 Jacc
- WordNet Jacc



内容が一致 ← 検索クエリで獲得した文同士のJaccard係数 → 内容がバラバラ

検索クエリで獲得した文と各拡張で獲得した文の類似度計算結果

# まとめ

- 完全に人手で普通名詞換言辞書を構築した  
換言対象の16,524語に対し、16,512語の換言対を得た
- 構築した辞書の評価としてクエリ拡張を行った  
普通名詞換言辞書と用言等換言辞書を合わせた換言辞書  
はWordNetと同等以上の効果があることがわかった
- 換言辞書は公開する予定